DOCUMENT RESUME

ED 426 082 TM 029 309

AUTHOR	Taylor, Terence R.
TITLE	Testing Approaches for the Future.
INSTITUTION	Human Sciences Research Council, Pretoria (South Africa).
REPORT NO	BIZ-2
PUB DATE	1990-00-00
NOTE	20p.; A summary in Afrikaans is included.
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Cultural Awareness; Disadvantaged Youth; *Equal Opportunities (Jobs); Foreign Countries; *Occupational Tests; *Personnel Management; Selection; *Test Bias; Test Use
IDENTIFIERS	*Fairness; *Learning Potential Tests; South Africa

ABSTRACT

Because applicant populations for many jobs and educational opportunities are becoming increasingly multicultural, the whole issue of fairness in testing has become highly salient. In South Africa, a main reason for the poor performance of the economy is the failure to make optimal use of all manpower, one of the crippling legacies of the apartheid era. There is also a strong moral motivation on the part of some employers to do something about the inequalities of the past by advancing greater numbers of disadvantaged individuals to skilled jobs. The question is one of how to identify the "right" individuals from disadvantaged backgrounds for these opportunities, since there are drawbacks to relying on conventional test scores to make selection, training, and advancement decisions. Three approaches to this problem are discussed in this paper: (1) putting in place of fairness programs; (2) use of trainability tests to make selection decisions; and (3) use of learning potential measures to make such decisions. Although practical learning potential measures suitable for application in industry do not yet exist, such measures are likely to be a useful tool. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 082

TM
PB98927

Report BIZ 2

1990

TESTING APPROACHES FOR THE FUTURE

Terence R Taylor

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

A. H. Stumpf

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM029309

Pretoria
Human Sciences Research Council
1990

PB98927

EXECUTIVE SUMMARY

Up until fairly recently the practice of psychological testing was not developing or changing very much. Political, legal, and social changes that have occurred over the past few years, however, have begun to have a significant impact on many facets of testing. Because applicant populations for many jobs and educational opportunities are becoming increasingly multicultural, the whole issue of fairness in testing is highly salient. This issue has been pushed into greater prominence by unions, who are beginning to question whether currently used selection practices, especially testing, afford their members a level playing field.

Most large employers also realize that the South African economy cannot continue underperforming unless we wish to become a full-blown member of the community of third world nations. A main reason for the poor performance of the economy is the failure to make optimal use of all manpower - one of the most crippling legacies of the apartheid era. There is also a strong moral motivation on the part of some employers to do something about the inequalities of the past and advance much greater numbers of disadvantaged individuals to skilled technical and managerial jobs.

The question is: how does one identify the "right" individuals from disadvantaged backgrounds for these opportunities? Although conventional test scores have their uses, they tend to be a record of the individual's disadvantagement. These scores are strongly correlated with socio-economic-status - a negative feature for this application. Therefore there are drawbacks to relying totally on these scores to make selection, training and advancement decisions, if the aim is progressively to give greater numbers of disadvantaged people with potential the chance to move up.

Three approaches to this problem are discussed in this report: the putting in place of fairness programmes, the use of trainability tests to make selection decisions, and the use of learning potential measures to make such decisions. Although practical learning potential measures suitable for application in industry do not yet exist, the author argues that these measures are likely to be a very useful tool to achieve the aims mentioned above.

OPSOMMING VIR HOOFBESTUUR

Tot taamlik onlangs nog was daar maar min vordering of verandering in die toepassing van sielkundige toetsing. Politieke, wetlike en sosiale veranderings het egter oor die afgelope paar jaar begin om 'n beduidende invloed op verskeie aspekte van toetsing uit te oefen. Omdat aansoekerpopulasies vir baie werks- en opvoedkundige geleenthede toenemend multikultureel van aard word, plaas dit die hele aangeleentheid van billikheid in toetsing op die voorgrond. Die aangeleentheid word verdere prominensie gegee deur vakbonde wat die regverdigheid van huidige keuringsmetodes, en veral van toetsing, begin bevraagteken.

Die meeste groot werkgewers besef ook dat die Suid-Afrikaanse ekonomie nie maar kan aanhou onderpresteer tensy ons van plan is om 'n volle lid van die derde wêreld te word nie. 'n Belangrike rede vir die swak vertoning van die ekonomie is dat nie alle beskikbare arbeid optimaal benut word nie - een van die mees verlammende gevolge van die apartheid era. Daar is ook 'n sterk morele motivering by sommige werkgewers om iets te doen aangaande die ongelykhede van die verlede en om groot getalle onbevoorregte mense na tegnies-geskoolde en bestuursposte te bevorder.

Die vraag onstaan nou: hoe identifiseer 'n mens die "regte" persone met ongunstige agtergronde vir hierdie geleenthede? Alhoewel konvensionele toetse hul gebruike het, is dit geneig om 'n rekord van iemand se onbevoorregte verlede weer te gee. Sulke toetse korreleer sterk met sosio-ekonomiese status - 'n negatiewe kenmerk in die lig van die beoogde aanwending. Daar is beslis nadele aan verbonde om algeheel op hierdie toetspunte aangewese te wees wanneer keurings-, opleidings- en bevorderingsbesluite geneem word, veral as die doel is om toenemend meer onbevoorregte mense met potensiaal die kans te gun om vooruit te beweeg.

Drie benaderings tot hierdie probleem word in hierdie verslag bespreek: die daarstel van billikheidsprogramme, die gebruik van opleibaarheidstoetse en die gebruik van leerpotensiaal-meetinstrumente om hierdie tipe besluite te neem. Alhoewel daar nog nie praktiese leerpotensiaal-meetinstrumente bestaan wat geskik is vir gebruik in die industrie nie, meen die skrywer dat sulke instrumente nuttig sal wees vir die bereiking van die genoemde doelwitte.

As I shall be covering quite a bit of ground in this report, it will be useful for me to give at the outset an outline of what I shall be saying. I shall be covering three main topics. Firstly I shall discuss some of the factors that, in my opinion, are changing the testing ballgame in this country. For many years the testing scene was very stable, almost moribund, but now things are changing rapidly and psychometrics people will, to coin and slightly modify a phrase from our previous esteemed State President, have to adapt or die out.

The second topic I shall be covering has to do with conventional psychometric tests. I shall discuss their strong points, the criticisms of bias and fairness that are levelled against them, and what I regard as some of their shortcomings, especially when applied in a country such as this, where there is such a wide cultural diversity and such huge disparities in opportunity - especially educational opportunity.

The third subject that I shall address concerns some possible responses that can be made to the problems and challenges that confront us. Here I shall look at some of the issues surrounding the implementation of fairness models; and I shall also look at alternative approaches to psychometric assessment, especially the learning potential approach, which I believe overcomes many of the criticisms of testing put forward by labour unions and certain other interested parties.

Developments in the Testing Ballgame

Let's go back to the first topic and look at some of the developments that are changing the testing situation in South Africa. The list that I have compiled is probably not comprehensive, but I think it contains most of the important factors.

1. Over the past decade or so the workplace has become increasingly integrated. Whereas previously there was job reservation and racial discrimination even when no legal strictures were in force, many job categories - especially at the artisan level - are now filled by individuals from all cultural and ethnic backgrounds. Previously selection tests were applied to members of only one group; now these tests, which were often standardized on a single group, have to be pressed into service to make cross-cultural selection decisions. As we all know, the ramifications of this are not trivial.

2. Political developments in the country are having repercussions for psychometric testing. Priorities are changing. It is my belief that the two

most important issues on the political agenda of the future will be (and to an extent already *are*) fairness of opportunity in the workplace and the equalization of opportunity and quality in education. Psychometric testing plays a prominent role in both of these domains.

3. As part of the changing power structure that is flowing from political developments, groups who were previously silent and relatively powerless are now making their displeasure known as regards certain practices in the workplace. These parties have not failed to notice the twin facts that the most desirable jobs are held by whites and that psychometric tests are commonly used to make selection and promotional decisions; as a result they have increasingly levelled criticism against the fairness of using these tests. It should be pointed out, however, that the critics of testing seldom suggest any workable alternative methods to apply in the selection process.

4. Up till about 1985 South Africa to a certain extent got away with - or appeared to get away with - having an economy that was performing at low levels of efficiency due to skills shortages. But with the collapse of the gold price and the value of the Rand, the ravages of apartheid and the effect that it has had on the economy is being felt by all, even the more affluent. Employers and government authorities are now fully aware that unless skilled manpower is drawn from a greater pool than the traditional male white one, the country is doomed to a steady decline into full third-world status. The need to identify people who have the potential to be developed, no matter what their ethnic and cultural background, is now widely seen to be very pressing. Tests can help in this regard; but unfortunately the tests that were used in the days of white male privilege (and are still being used) might not in some cases be suitable to identify the best person - irrespective of group membership - for the job or educational opportunity.

5. There has been pressure from overseas, especially on companies with foreign principals, for fairer employment and promotion practices. Several codes of fair practice have been framed, some of these being of local provenance, and many companies have adopted one or other of them. Selection and placement testing are very important components of employment practice. Low standards were common with regard to the types of tests used, the test administration practices applied, the quality of decisions made from test scores, the validation of tests, and the checking of tests for their suitability in cross-cultural selection contexts. The codes put pressure on test users to revise these standards upwards.

6. Apart from pressure from outside our borders, there is a genuine sentiment on the part of many large employers that the employment practices traditionally applied in South Africa are ethically indefensible. Many employers are now looking for ways of making amends and helping to address the disadvantagement which is the lot of so many people in this country. These employers want to identify those individuals in disadvantaged communities who have potential and give them opportunities, such as financial support for further studies. Clearly tests can play a role in this process, but not always the tests that have been traditionally used.

Conventional Tests: What we Can and Can't Rely on them to Do

Whatever happens in this country, there are still going to be more people wanting desirable opportunities (such as jobs and places in educational institutions) than there are opportunities. Therefore some sort of selection procedures will continue to be needed, and the more objective these are, the better. Tests, ironically, are often criticized exactly because they are, to a large extent, objective tools: their very objectivity allows their flaws to be detected. Really "soft" methods - such as the interview - often cannot be empirically shown to be inadequate because they produce no objective output.

As tests come under a critical spotlight we should not forget that they have a number of positive features:

* They are available in some variety;
* They have been extensively used and researched, and hence their performance is fairly well understood;
* They have shown reasonably good validity over a wide spectrum of criteria;
* And, as I stated earlier, they produce objective results.

Against these positive features we have to weigh the criticism that tests are biased against people who do not have a western or "white" cultural background. In America these criticisms have often been found not to have substance when statistical evidence is collected, but we should remember that South Africa is not the USA. For a start, different cultural groups speak different languages here, whereas English is the dominant language in America, and the most widely researched minority group - Afro-Americans - all speak English.

There are three main types of bias or comparability: construct, item, and predictive. Construct comparability has to do with whether a test measures the same psychological dimension in different cultural groups. In order to investigate construct comparability, one has to examine the pattern of correlations of the test in question with a variety of other measures and indices. If the pattern differs across cultures, evidence of construct incomparability exists.

Item comparability or bias is concerned with phenomena at a more micro level. Item bias investigations ask the following question: "Do members of a given group have more difficulty getting a particular test item right for reasons that are unrelated to the ability in question? If a test of intelligence contains items on English proverbs, then these items are likely to be biased against second-language English speakers. Such items are particularly undesirable because not only are they inappropriate as measures of intelligence for second-language speakers, but they are also inappropriate as measures of language competence in the work situation. (How often is it critical to one's job performance to know what "The burnt child dreads the fire" means?) Obviously it is the test publisher's responsibility to keep biased items out of new tests and to address the problem of item bias in older ones. The NIPR is busy with these tasks at present.

For practical test users the most important kind of comparability or bias is predictive bias. Predictive bias investigations attempt to answer the question: "Do test scores translate into the same level of criterion performance in different groups?" Clearly problems arise with regard to fairness if the answer to this question is "no" and the same test cutoff is used for different groups.

This brings us to the issue of the relationship between fairness and bias. Many people assume that if a test is shown to be biased (e.g., to have some items that are biased), the test can be referred to as "unfair." This conclusion is not meaningful - because tests per se cannot be referred to as fair or unfair. Fairness rests in the use to which a test is put, not in the test itself. A biased test can be used fairly and an unbiased test can be used unfairly.

The concept of fairness is not absolute. Like so many things in this life, fairness is not some untouchable Platonic Idea but a construct devised and measured out by human minds. One man's fair is another man's foul. The conception that individuals or parties develop with regard to fairness is very much conditioned by their role in the world of work or education. If you are looking for a job or a place in an educational institution, your idea of fair selection is likely to be very different from that of the party on the other side of the fence: the employer or registrar.

There are almost as many models of fairness as there are people who have thought about this issue. However the models that have been proposed in the literature fall fairly easily into three classes. The first class contains those models that are based on quotas. In these models, members of different groups are hired according to certain proportional rules. For instance, an organization might strive to hire members of different groups in the same proportion that they occur in the applicant population.

The second class of fairness models is usually known under the name of unqualified individualism. Models that fall in this category are based on the view that any information that improves prediction of the criterion can and should be used in the selection decision. Hence group membership is a valid piece of information to take into account when deciding whether to offer a person a job. If members of a certain group have shown themselves to be more likely to do well on a particular criterion than members of another group, then preference is given to members of the former group.

To some fairness commentators, the practice of using group membership as a predictor is unethical, possibly even repulsive. They see the unqualified individualism approach as leading to the entrenchment of current inequalities of opportunity and perpetuating social problems. People of this view often also see the quota approach as unacceptably group conscious. They tend to be attracted to the third class of fairness models which are called qualified individualism models. The qualified individualism models are the only ones that are truly "colour-blind" and "sex-blind". Like the unqualified individualism models, these are concerned with optimizing the probability that candidates will be successful on the criterion - but within the constraint of disallowing any information that refers to group membership, either directly or indirectly.

Hybrid approaches are also common. For instance, group membership might initially be taken into account, but only to find a subset of predictors that perform similarly on different groups. Thereafter, a group-blind approach is adopted. This is called a procedure of sub-optimization.

We must now terminate this diversion into some of the main concepts of fairness in testing and return to a discussion of the nature of conventional psychological tests. I mentioned earlier that these tests have a number of strong points; but they also have some shortcomings. Three of these seem particularly serious to me. The first of these concerns the grossness of the constructs measured. Conventional tests tend to measure very broad psychological characteristics, such as "verbal ability" or "general intelligence". While these constructs have their uses, especially

in making selection decisions in societies where there is a large degree of homogeneity, they also fail to tell us many valuable things about people. For instance, they do not tell us why certain people or groups of people tend to do poorly. A low score can be for a variety of reasons; but conventional tests are not designed to pinpoint causes of poor performance: in other words they are not diagnostic. In a society, such as ours, where many people are striving to be competent in a cultural milieu that is not their native one, it is very important that tests pinpoint specific areas of difficulty that individuals may be experiencing. This information can be used to design suitable educational, training, or remedial packages and possibly even to restructure jobs.

The second shortcoming of conventional tests is that they tend to be rather culture-loaded. The item material tends to contain references to certain phenomena that are familiar to one culture (the culture of the test constructor) but less familiar to other cultures. Obviously this is a serious problem in a multicultural society such as the South African one. The cultural specificity of test material is the main underlying causal factor of test bias.

In my opinion, the third shortcoming of conventional tests, the one that I shall describe now, is the most serious. This is that conventional tests "look back" rather than "look forwards". Conventional tests assess the individual's repertoire of strategies and knowledge. As disadvantaged people tend to have fewer of these due to their limited opportunities, they also tend to score lower on conventional tests. Hence conventional tests scores are to a degree a "record of disadvantagement". This is evidenced by the fact that there is a strong correlation between socio-economic status and test scores.

In order to redress the inequalities of the past, we need to find ways of evaluating an individual's potential rather than confirming his disadvantagement. This is the forward-looking approach. It is true that we cannot wipe out the effects on ability of past discrimination and poverty of opportunity by adopting this approach. Past disadvantagement lives on in the present in the form of reduced effectiveness in the workplace and in the educational institution. But we can give those who show, despite their disadvantagement, evidence of potential, the opportunities for development and advancement. Conventional test scores are not good indices of potential; and hence selection decisions taken on the basis of conventional test scores largely ignore potential and concentrate on current performance. As with the unqualified individualism models, we have here a case of the perpetuation of the inequalities of the past and the social and industrial problems that attend this.

We must find answers to the problems of selection that beset us and find these answers rather fast. Political and social developments are occurring so rapidly at present that we have at most only a few years to put the testing house in order or risk having the whole practice of testing swept away. This would be a great pity because tests - the right kind of tests applied properly and with the test scores used responsibly - offer the best chance of doing selection and personnel development in a "fair" and objective way. I place "fair" in inverted commas because there is no one fair way to do selection; but one must be true to the precepts of the fairness approach that one adopts: one must live by one's stated policy.

As tests give objective quantified outputs, their responsible use increases the chances that fairness policies have real and positive outcomes "on the ground". In many cases, good intentions by top management and enlightened policies have little practical outcome. Often the reason for this is that soft selection tools and vague application rules are in place, which be can "fudged" by those lower down - who possibly have an interest in maintaining the status quo. This fudging is less easy to do when quantified test results are used to make selection decisions.

The Road Ahead: Three Possible Responses to our Problems

I want to speak about three courses of action that can be taken to address our current problems. The first has to do with implementing fairness programs, especially the aspects that are relevant to testing. Then I wish to mention trainability testing as a possible way of overcoming some of the problems of conventional tests. And finally I shall elaborate on the theme that I touched on earlier and discuss the design of "forward-looking" or learning potential tests and their role in the future South Africa.

The three courses of action that I shall mention here are certainly not the only ones. For instance I omit to speak of the theoretical and practical issues involved in creating truly diagnostic, "information processing", psychological measures that can be used to assist designers of training material and remedial programs. That is a large topic in its own right - and one where the problems of turning theory into usable instrumentation are by no means trivial; I shall not touch on it here for fear of doing it injustice by having to be too brief.

Implementing fairness procedures

At the outset I want to make it clear that "fair" personnel practices embrace much more than simply the ethical and responsible application

of tests. Fairness issues come into play even before a potential employee steps through the door of the organization's premises: it begins with recruitment practices. Does the organization make its vacancies known to all groups of potential employees or only to the "good old boys", the members of the "club" who have traditionally held jobs - or at least the better jobs - in the organization? And fairness issues do not "go away" even after the individual has retired from the company. Certain areas stand out as being particularly important as regards fairness: these include recruitment, selection, and manpower development (including training, career path and advancement opportunities).

Testing is mainly relevant to the selection aspect of manpower utilization, but test results are also sometimes used (not always justifiably) to make decisions regarding placement, promotion, and training opportunities. Organizations have to think carefully about what role they wish tests to play in these and other domains; and they should investigate whether the tests they use are indeed providing valid information for the applications in question. Most personnel practitioners are aware of the importance of doing this in the selection context, but tend to be lax about it when using tests for making decisions about the person once he or she is accepted into the organization. The issues of across-group fairness are relevant to all applications of tests.

In my opinion there is no one right way to set out the steps required for installing a fairness programme, but the following are important elements that will have to be incorporated into such a program:

* The framing of a policy statement on the organization's stand on fairness as it relates to employment and treatment in the workplace;

* The casting of this policy into a precise model of fairness;

* The clear statement of the fairness goals towards which the organization will strive, preferably with target dates listed;

* Working out of the detail of implementation (taking a "big" view of what fairness embraces and not just looking at tests) and evaluating the feasibility of the envisaged procedures;

* Undertaking research into the performance (especially across groups) of the selection procedures that are currently in use and of other procedures whose application is being considered;

* The putting into place of the actual mechanisms and infrastructure required to implement the fairness procedures;

* Continuous monitoring of the effectiveness of the procedures, the comparison of actual achievements with stated goals, and the implementation of any corrective steps that seem necessary in order to attain the goals.

The NIPR has publications that discuss these issues in some detail.

Trainability and work-sample tests

All large and medium organizations should go through a set of steps similar to those listed above in order to improve the fairness of their personnel practices. But currently available tests might make it difficult to achieve goals that the organization sets for itself. What does the company do, for instance, if it wishes to increase its representation of blacks in particular categories of jobs to a certain proportion but finds that the distribution of test scores of this group (especially on verbally-based measures) simply makes it impossible to select enough people? Different cutoffs can be set, but this course of action is problematic for a variety of reasons.

If it is found that many of the individuals of a particular group fail to reach acceptably high scores on a test or battery of tests, but that - given the chance - they do better than expected on a criterion (such as a training course), then it is necessary to consider different selection instruments, possibly ones that are more closely related to the criterion in question.

It is in situations like this that trainability and work-sample tests can have their uses. These tests contain material drawn from the domain of the criterion; usually the process of selecting it is undertaken in consultation with skilled practitioners of the activity in question. The material is then turned into a lesson, after which the testees do whatever they have been taught. An attempt is made to design the task in such a way that assessment is highly objective. As both psychometric and technical considerations are important, these tests are quite tricky to construct.

Trainability tests have several advantages:

* As they evaluate what the person acquires after a lesson on the activity in question, they can be particularly suitable for selecting people for training programs;

* They can have very good predictive validity for the criteria for which they were designed, especially when the prediction is done in a relatively short time frame;

* They are more immune than conventional tests from attack by anti-testing critics because they seem more relevant to the criterion and hence are regarded as less biased;

* Job applicants tend to accept them without question.

However, these tests also have certain disadvantages:

* They have to be updated as jobs or training programs change;

* Their validity is suspect when used for selecting for criteria other than those for which they were designed (the range of criteria over which they are predictive might be rather narrower than is the case for conventional tests);

* They can be expensive to produce, especially if many have to be constructed for different criteria;

* Administrators of the tests have to have both technical and psychometric expertise;

* They are not really suitable for giving a long-term and broad picture of a person's potential: this presents a problem if the evaluator wishes to map out a career for an individual rather than make short-range decisions about his or her future.

Learning potential tests

Learning potential is a concept that originated with the Russian psychologist Vygotsky and was further elaborated by the Israeli psychologist Feuerstein. Vygotsky developed the idea of the zone of proximal development which is the degree to which a person can improve his or her performance after a session of intervention and help from a more skilled person. Feuerstein was particularly concerned with diagnosing cognitive lacunae in adolescent immigrants to Israel who were experiencing difficulties in high school. He developed the Learning Potential Assessment Device (LPAD) which permits the assessor to

determine where remediation is necessary and what cognitive growth potential the person has. The LPAD is coupled with the Instrumental Enrichment (IE) device which is used to address deficits.

Feuerstein applies very flexible, but not particularly standardized, procedures. In the LPAD, the individual is assessed using a variety of cognitive tasks (such as a version of the Ravens Matrices). Then he is given a variety of exercises designed to help him develop the skills required to solve problems that he was previously incapable of doing. Finally he is reassessed on the tests that he was given in the first part of the exercise. The difference in his performance is taken as his learning potential.

In practice, this difference is apparently seldom quantified. Feuerstein's approach is rather clinical. It requires highly skilled personnel to administer the measuring instruments and perform the interventions. And, being a one-on-one procedure, it demands vast amounts of skilled time. Clearly the LPAD approach is inappropriate for industrial selection and placement applications.

Our first attempt to overcome these difficulties involved the development of a test-teach-test technique which could be administered to groups and which produced quantified difference scores. Unlike the LPAD the subject was confronted with only one type of task - a set of letter series problems. Also unlike the LPAD, the pretest and posttest were not identical: rather they were parallel forms of the letter series task. The lesson that the subjects did after the pretest was administered by the tester, but was also made available to subjects in written form. They had to do various exercises during the course of the lesson.

A serious problem that arose with this method was that the strategies that the subjects had developed in the pretest often clashed with the methods taught during the lesson. As a consequence, a sizeable proportion of subjects actually got worse rather than better in the posttest. Despite this shortcoming, the difference scores (between the pretest and posttest) correlated significantly with educational criteria. They also correlated almost not at all with conventional ability scores; this indicates that the difference score was giving information on testees which cannot be garnered from conventional test scores.

It is my opinion that the best way to assess learning potential in the industrial selection context is to do away with the lesson and replace the pretest and posttest with a larger number of very short tests. In the initial instructions the subject is introduced to the basic nature of the task and shown how to do it. The task should be of such a nature that everyone can do it after receiving the instructions and practising on a few practice items. But they should be slow and inefficient at doing it. In order to

PB

prevent people from becoming proficient at it in only a few minutes, the task should have multiple steps and require information to be accessed from one or more sources (which could take the form of separate sheets of paper or even a small booklet).

Once the testees know how to do the task (but not how to do it efficiently), they are given a series of short time-slots in which to do as many units of work as they can. These time slots should be in the vicinity of 3 or 4 minutes. After each session the subjects are asked to mark where they are; there is a brief pause and then they are started on the next minitest. Altogether there might be 8 or 10 minitests.

As individuals receive no actual coaching on how to do the items efficiently, improvements in speed and accuracy are a result of "getting one's act together". The well-known cognitive psychologist Sternberg calls this process "automatization" and regards it as a very important aspect of intelligence. Automatization reflects metacognitive activities -the cognitive activities that direct other, lower level, steps of problem solving. Most people can do the lower level steps, but not everyone can put these steps together to do a task in an optimal way. Automatization therefore results from a process of self-instruction whereby the individual uses his executive or metacognitive skills in order to put together a program of lower order processes to do a novel task more and more efficiently.

Automatization is important partly because the faster one becomes automatized, the sooner one has mental capacity available to tackle the next novel task. People who automatize quickly tend to learn more and therefore become more skilled over a wide range of domains. (But of course this building up of repertoires of skills is moderated by the opportunities that one has.) Automatization and response to novelty are, according to Sternberg, two sides of the same coin.

Response to novelty (and cognitive flexibility) can be studied by giving subjects a second task after they have had the 30 or 40 minutes experience on the first. This second task should have some features in common with the first but also some aspects that are different. The faster the subject can modify his existing "program", the faster he or she will become proficient at the second task. Individuals who have the flexibility and metacognitive skills required to modify their cognitive "programs" effectively should become proficient at the second task much faster than they become proficient at the first. Cognitively rigid individuals, on the other hand, are likely to have difficulty disassembling and modifying their "old" programs; hence their learning curve will be less steep.

A final step in the assessment exercise can be added. This involves testing how much factual material the subject has acquired in the process of doing first and second exercises. While doing these tasks the subject

has to use information which is located in the separate sheets or booklet but is not required to memorize this information. More effective performers will, however, try to remember as much as possible because this will cut down on the amount of reference work that has to be done. The amount of information that the individual has acquired can be assessed by removing the loose sheets or booklet from him and testing him on his knowledge.

This three-step approach to cognitive assessment permits one to get a handle on the actual dynamics of knowledge and skill acquisition. As I mentioned earlier, the evaluation of learning potential - of the acquisition of factual and procedural knowledge - gives us a new window on the person, a window that does not just provide us with another view of his disadvantagement. Conventional ability tests, on the other hand are to a large extent a record of the person's privation.

The dynamic picture that this approach gives of the individual's cognitive functioning is illustrated in Figure 1 which shows the subject's performance against the performance of the norm group. Both the reference group's and the individual's performance are normed. Standard scores are used in this illustration (which have a mean of 50 and a standard deviation of 10), but any other norming procedure would also be acceptable. The sessions (10 in this case) are ranged along the horizontal axis. By definition, the norm group average for each session is 50. Although learning is obviously occurring, the norming process returns the average value of the standard score to 50. The subject's performance can easily be compared with the average performance of the norm group simply by looking at his or her

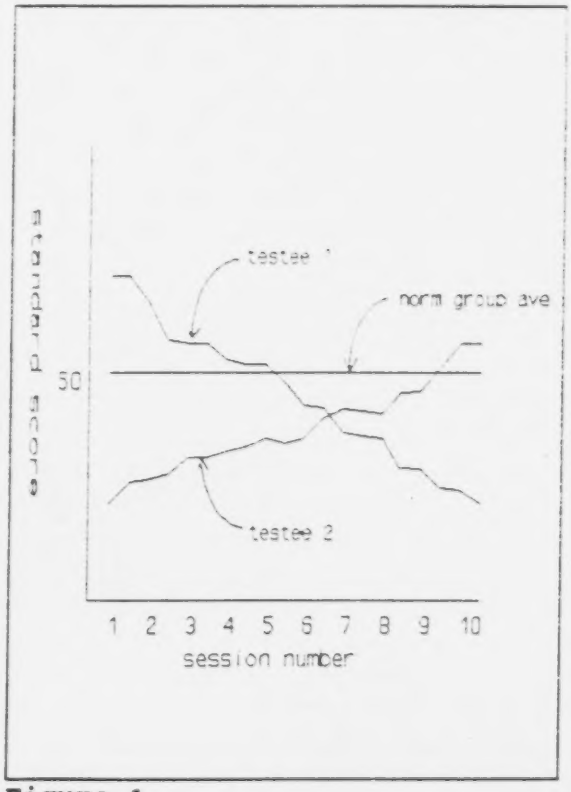

Figure 1

normed score against the 50 for the norm group.

If there is a general upward trend then the subject is learning faster than the standardization group and the opposite is true if there is a downward trend. If the subject's line remains below the norm group's line throughout the session but is steadily moving upwards, then the individual, although below average in his performance on all sessions included in the test, might still be worth considering for various jobs or training courses, as long as the pace of learning required on these criterion activities is not too rigorous. If a subject's line starts above the norm line but then declines and possibly even crosses below the norm line, the individual might be suspected of lacking staying power. Many other interpretations can be drawn, depending on the shape and position of the subject's line relative to the norm group's line.

Interesting results can also be obtained from the un-normed data. Figure 2 reflects the subject's performance and the norm group's performance in actual work units per session. An exponential curve has been fitted through each set of points. One of the parameters of the exponential curve reflects the actual learning rate. Hence it is possible to compare, in a single score, the individual's learning rate against that of the norm group. In addition, the average deviation of the subject's scores from the exponential line might be an informative index of his or her consistency in acquiring new skills.

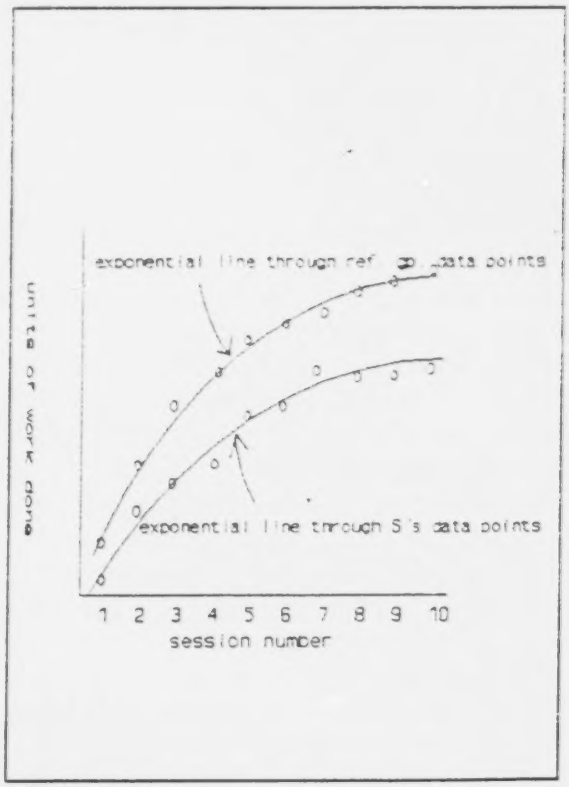

Figure 2

Up to this point we have been looking at the results only of the first learning exercise. However we can also compare the individual's and the norm group's learning curves across the first and second exercises. Individuals who respond to novelty through

a swift restructuring of their strategic program will tend to have appreciably steeper learning curves in the second session and this will be reflected in the learning rate parameter. The difference in an individual's learning rate between the two sessions can be compared with the norm group's difference. All these comparisons are informative of the person's learning potential.

To summarize, then, the learning potential approach of the kind sketched above seems to have the following advantages:

* It "looks forward" and measures future potential rather than past disadvantagement;

* It should give more general information on the person's capacities than trainability tests;

* It constitutes a creative solution to the fairness problems inherent in the use of conventional tests;

* Unions are likely to find tests based on the learning potential approach acceptable as selection tools.

Studies done in Africa have shown that people to whom testing is a foreign concept do much better if they receive careful comprehensive instructions and if they are given ample practice before starting the test. In learning potential tests of the kind described above, the very process of acquiring familiarity with test material is an integral part of the measuring procedure. Although the task required in the learning potential exercise is new to all testees, some might find the whole exercise more unfamiliar than others. But they still have the opportunity to show their mettle by the way in which they come to terms with the material and become more proficient at handling it over time. Conventional tests do not measure this dynamic aspect of performance and hence do not tell the whole story of a person's capacities: they tell of capacity but not capacity to attain capacity.

I do not mean it to be concluded from this comment that conventional tests are useless. They have their place. Current level of ability is a very important indicator of a person's skills that he or she can implement on the job or training course *right now*. This information, of course, cannot be ignored. But it should be considered in conjunction with one or more measures of future potential.

In the new South Africa we shall try to redress the discrimination of the past and companies will no longer be able, for reasons of ethics and

manpower availability, to take in only those individuals who are ready to do the job straight away. Greater responsibilities will rest on organizations to put in, in the form of training and job opportunities, what the formal educational process left out. It will fall to big companies to pick up much of the tab for the "sins of the fathers". I believe that learning potential measures will be a useful tool for them to use in this process.